MILFORD HAVEN: ENERGY KINGDOM

Data Ecosystem report

Prospering from the Energy Revolution Data Ecosystem Strategy

March 2022

ENERGY KINGDOM
DEYRNAS YNNI

Table of contents

Executive summary

The Milford Haven Energy Kingdom (MH:EK) project aims to become a Smart Local Energy System (SLES) with leading innovations towards a decarbonised energy future. Part of the project has involved carrying out detailed energy modelling to determine the most optimal solutions and propositions to seek investment in.

This energy modelling relies on a broad variety of data in order to determine the impact of changing supply and demand in the energy markets as a result of the decarbonisation propositions. The quality of data and its management will be critical to the success of such initiatives, and this report sets out the key areas to focus on within the data lifecycle from the specification of data to the core enablers such as people, process and technology.

This study assessed the use and management of data throughout the project and posed the question of whether the experience found on this project matched what had been observed in the Energy Systems Catapult Data Taskforce report (published in 2019) *[5].* This outlined some of the challenges faced by the sector regarding data sharing and management, making some recommendations to improve the situation.

Key findings

This study compared the project with the recommendations of the Taskforce and found that in general a similar picture was found at this more localised level. It was however felt that the ability to control and influence data standards and requirements was limited for those organisations providing the data for the modelling exercise.

The key actions which the MH:EK SLES group could make were largely limited to improved documentation of the modelling outputs, techniques and publication of any manipulated datasets where data agreements allow.

.

Recommendations

There are a few key national energy sector initiatives which are underway such as Open Energy *[10]*, Virtual Energy System *[9]* and Future of Gas *[14]* which will enable a much better integration of MH:EK SLES into the wider energy market through better data sharing and standardisation. These initiatives however require the representation of local systems such as MH:EK to ensure that their needs and capacity are considered. Therefore, it is recommended that where possible representation in these advisory groups be sought.

The main recommendation for the MH:EK project is that it has plans in place to prepare for initiatives such as open data, standards and a focus on the fact that having available and accurate data will be to its advantage when some of the outcomes from the national initiatives become a reality. Throughout the lifecycle of the design, construction and operation of the propositions, the data required from these assets for their maintenance, and for the wider energy sector will be required as part of the delivery.

It is recommended that a data working group be established within the MH:EK organisations in future to ensure that the various data initiatives recommended in this report, and within the energy sector, are discussed and championed locally in a coordinated way.

.

Purpose of this document

Section	Operation
1. Overview of Project	An introduction to the sector outcomes and drivers which could be partially met through data enabled decision making and the energy modelling on the project.
2. Current State of Energy Data	An assessment and analysis of the project processes and the data that enables them. This includes the solutions that have been implemented alongside the challenges that have been encountered from data specification through to use.
3. Future State of Energy Data	Consideration of the future of both energy data and modelling across the sector including a future data management vision alongside the ideal future state for the MH:EK project.
4. Recommendations & Roadmap	Recommendations for the improvement of processes and data across MH:EK with comparison to sector-wide recommendations.
5. Conclusions	A summary of the key findings from the report.

Table 1: Overview of this report.

The Milford Haven: Energy Kingdom (MH:EK) data ecosystem is required to ensure the existence of the data lasts beyond the initial 2-year project. A data ecosystem should contain all project information and data throughout the lifecycle from design through to operation and ultimately the demolition of assets within the energy system.

The energy sector is currently embarking on a journey of 'A Modern, Digitalised Energy System', and has set a series of steps to achieve this. This system is key to decarbonise the energy sector, and data management is one of the main enablers that will assist the sector in reaching its goals.

The MH:EK project has involved liaising with multiple data owners and data providers to assess, review, and draw insights from multiple datasets to inform the Smart Local Energy System (SLES) design.

The project is one of the ten InnovateUK funded 'Prospering from the Energy Revolution' (PfER) detailed design projects, which intends to develop SLES's that intelligently link the energy sector vectors of power, heat and transport towards zero carbon.

This document aims to define and describe:

- The Energy Sector objectives to enable decarbonisation and decentralisation
- How these objectives will be achieved through effective decision making enabled by data and modelling
- The lifecycle of energy data which will deliver the information for the decision making
- The current situation of data within the sector as experienced on the MH:EK project
- A review of the Energy Data Taskforce report ' A strategy for a Modern Digitalised Energy System', specifically challenges and recommendations
- A roadmap to determine the next steps for the sector, as well as the MH:EK project to best align it with the sector and its objectives.

Document guidance

A robust data management process is required as part of the proposed data ecosystem. This enables data driven decision making whilst considering the project drivers, which results in the delivery of the outcomes of the sector.

Figure 1 below shows the energy data lifecycle that will be used within this report to assess and analyse the data that has been collected and implemented on the MH:EK project.

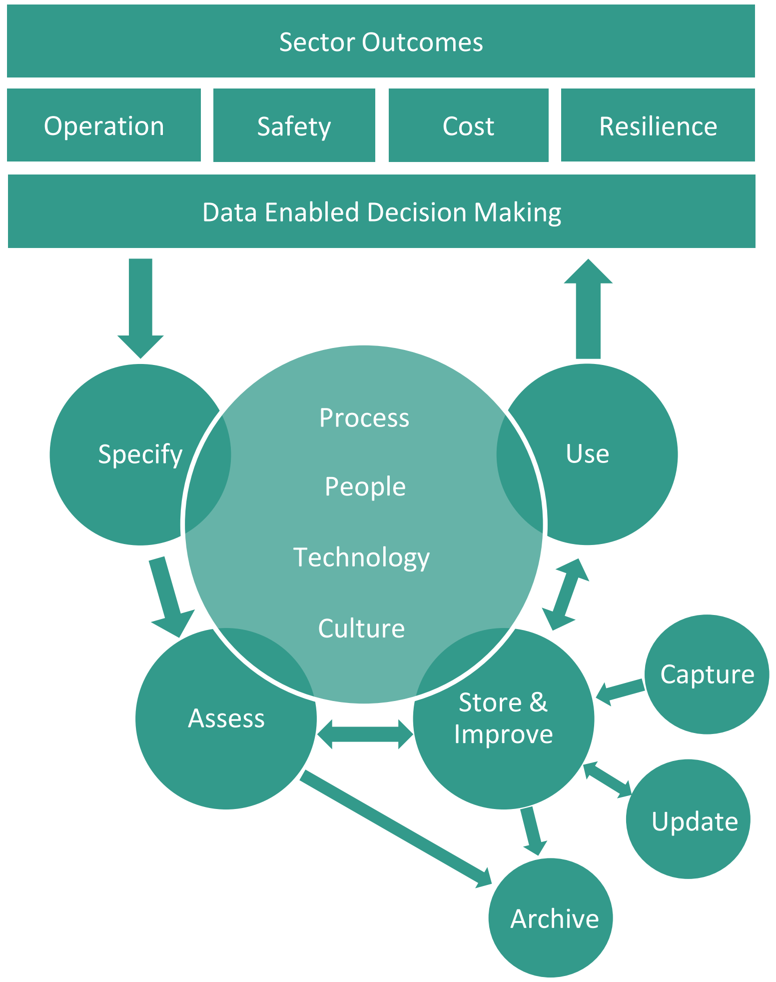
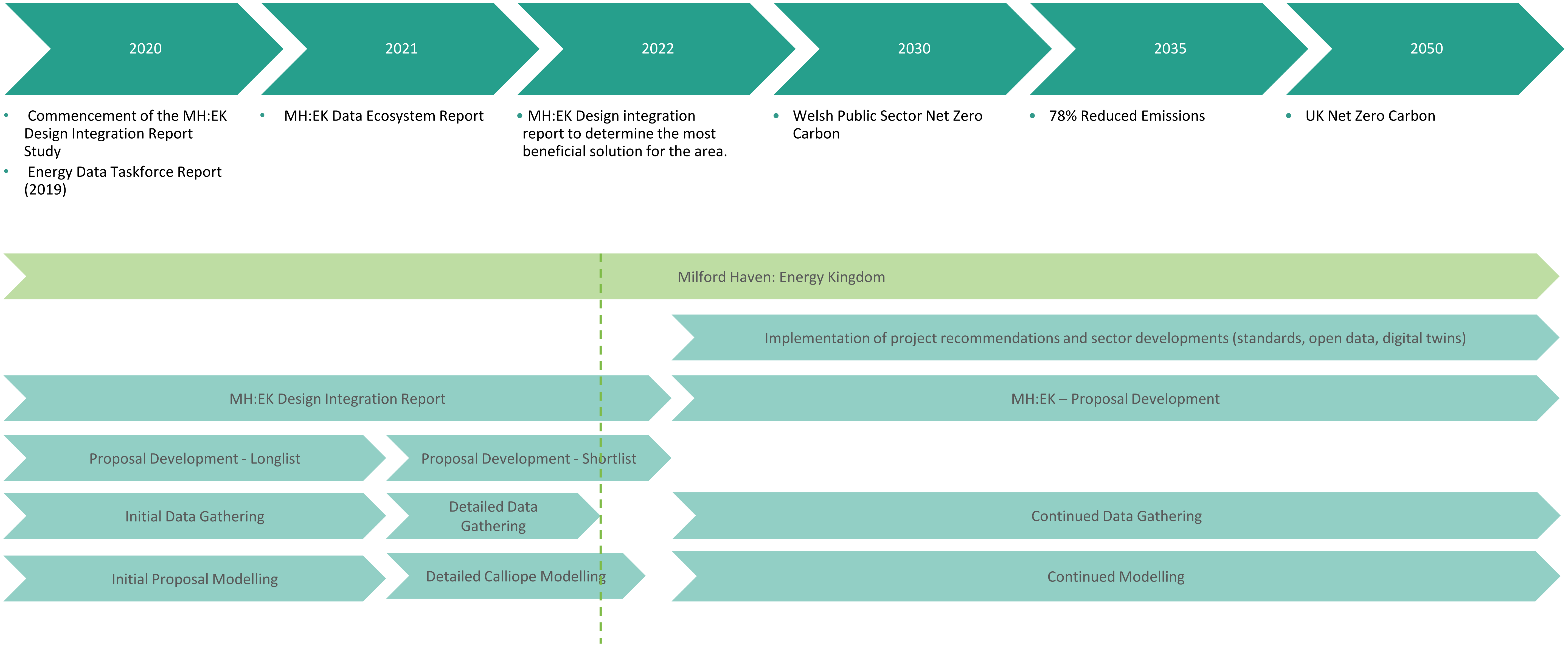


Figure 1 – Energy Data Lifecycle

Project timeline

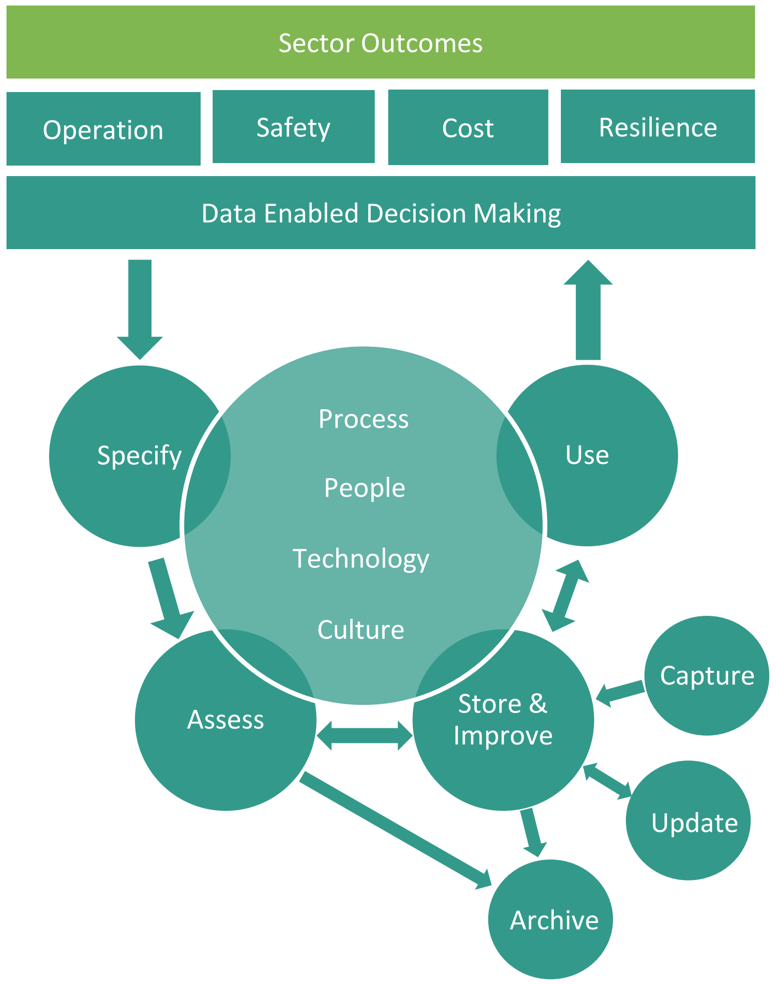


1. Overview of the project

The data and modelling outcomes and process

ENERGY KINGDOM
DEYRNAS YNNI

Energy data lifecycle: Sector outcomes



Sector outcomes

In June 2019, the UK Government set a target, and passed laws, that will require all greenhouse gas emissions to be net zero by 2050 [1]. In April 2021, the UK government set a further target to reduce emissions by 78% by 2035 compared to 1990 levels [2].

The Energy Systems Catapult (ESC) was set up in 2015 to accelerate the transformation of the UK's energy system. The Energy Data Taskforce (EDTF), run by the ESC, was commissioned by the UK Government, Ofgem and Innovate UK to develop an integrated data and digital strategy that helps unlock the opportunities of a modern, decarbonised and decentralised Energy System for the benefit of consumers [3].

In June 2019, the EDTF released the report titled *'A Strategy for a Modern Digitalised Energy System' [4]*, which provided a five staged approach to achieve a Modern, Digitalised Energy System:

- **Data Visibility:** Understanding the data that exists, the data that is missing, which datasets are important, and making it easier to access and understand data.
- **Infrastructure and Asset Visibility:** Revealing system assets and infrastructure, where they are located and their capabilities, to inform system planning and management.
- **Operational Optimisation:** Enabling operational data to be layered across the assets to support system optimisation and facilitating multiple actors to participate at all levels across the system.
- **Open Markets:** Achieving much better price discovery, through unlocking new markets, informed by time, location and service value data.
- **Agile Regulation:** Enabling regulators to adopt a much more agile and risk reflective approach to regulation of the sector, by giving them access to more data of higher quality and granularity.

As MH:EK is producing a design for a Smart Local Energy System (SLES), and the process is looking to be replicated across the UK for other projects, the sector outcomes have been at the forefront of the project and will also need to be at the forefront of the data ecosystem as the project progresses.

The ESC Data Taskforce's staged approach in Figure 2 is a process put in place to decarbonise the sector, many of these stages may take many years to be effectively implemented. The MH:EK data ecosystem and project data management systems and processes will need to be in place to enable its incorporation as the sector progresses.

This report will outline the data challenges that were faced on the project to provide a more detailed example of likely issues faced in the wider sector. The lifecycle of the data and supporting information required for energy modelling will be presented, and a roadmap of future actions and recommendations will be provided.

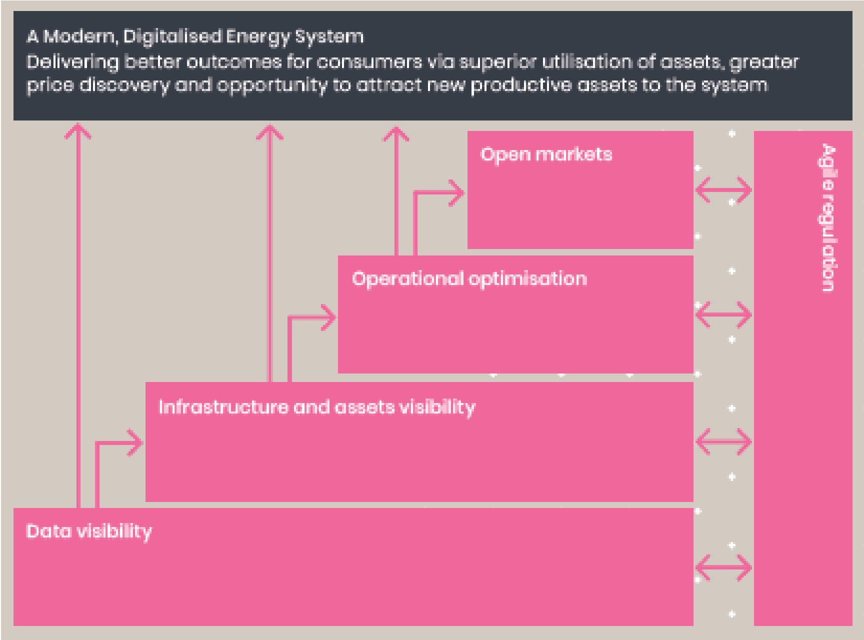
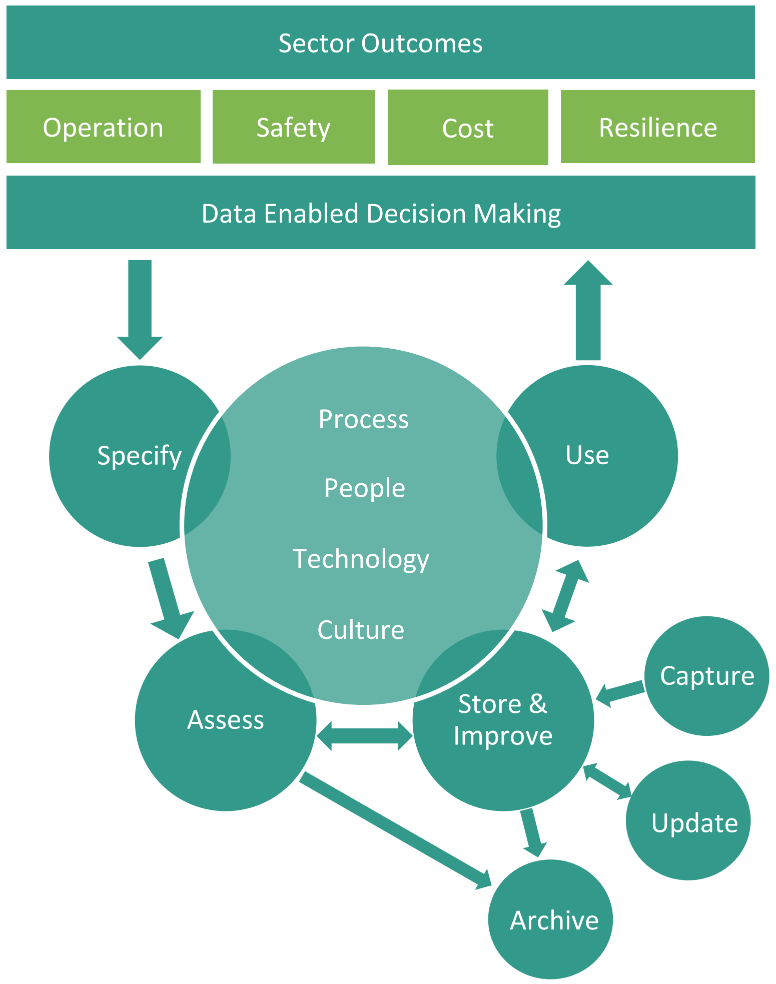


Figure 2 – Visualisation of the five stages highlighted to achieve a Modern, Digitalised Energy System [4]

Energy data lifecycle: Sector drivers



Sector drivers

To achieve the sector outcome of a decarbonised and decentralised energy system, there need to be sector drivers that contribute to the overall outcomes. These drivers are achieved in part by the data driven decision making on projects. The following sector drivers have been highlighted as part of the MH:EK project and have been used to assist and inform decisions on project direction and outcomes.

Operation

The selected proposition must be operational in the context of the existing infrastructure, or infrastructure that can reasonably be created, maintained and operated in the region. Its operation must also meet the local skills and infrastructure available in the region and meet with regulatory requirements.

Safety

Any proposition must meet all health and safety requirements, ensuring it is at the forefront of the decision-making process. This must be considered throughout the lifecycle of any project from planning through to operation.

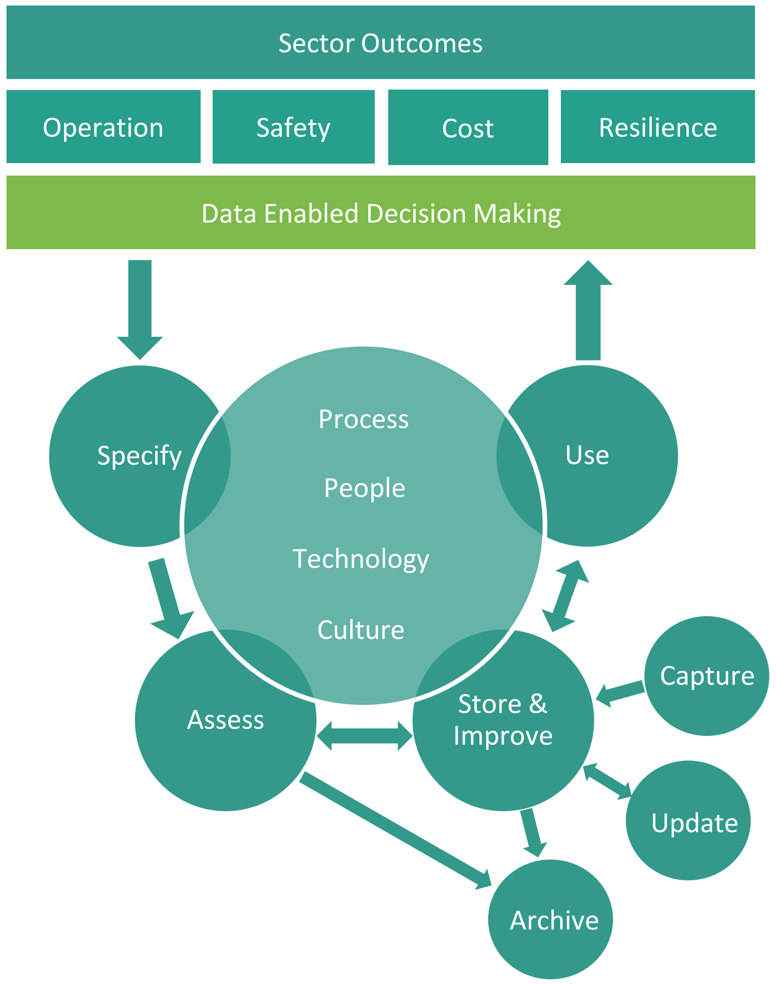
Cost

The best proposition from an operational / technical perspective could potentially be at a very high cost either of construction or operation. A proposition must therefore be commercially viable and compete reasonably with other energy alternatives and investments.

Resilience

The propositions must be tested under multiple possible scenarios. This will allow comparison of technologies and solutions to ensure the best and most resilient solutions are proposed and progressed. Enabling low-regret decisions to be made in the immediate to short-term that support alternative future scenarios out towards 2050+.

Energy data lifecycle: Data enabled decision making



Data enabled decision making

The project looks to answer key questions on renewable energy, using data within an energy modelling process to determine the most effective propositions to be implemented. Provided below against each sector driver are some of the questions which the MH:EK project is looking to answer, many of which can be determined via the energy modelling and data visualisation activities being undertaken.

Operation

- What are the best energy solutions to decarbonise the energy sector?
- What is the Milford Haven delivery roadmap to net zero?
- Is hydrogen beneficial for consumers?
- Which hydrogen solutions are the most effective?
- How is hydrogen best integrated into the energy system to decarbonise energy supply?
- Are consumers ready to adopt proposals even if the benefit is demonstrated?
- What does "best solution" mean, and how do you demonstrate it?
- Can hydrogen be used to alleviate renewable curtailment and network pressures – meaning investment in the network can be reconsidered?

Safety

- What are the safest technologies to construct, implement and maintain for the Milford Haven energy system?

Cost

- What carbon price is needed to make hydrogen a viable energy vector?
- What are the recommendations for wider infrastructure investment that support the ongoing MH:EK energy system development?
- How do Government incentives and market signals drive investment in hydrogen infrastructure?
- How can, and what value, of energy bill or vehicle maintenance costs could be impacted by a switch to hydrogen?

Resilience

- How is it ensured that the proposed solutions will be resilient in the future and new technologies can be continually incorporated?
- How is it ensured that the proposed system is resilient and can supply sufficient energy to the local area?
- How will the use of both existing and proposed assets be optimised?
- How can a long-term job/skills pipeline be ensured in the MH:EK area to support the wider UK energy system?
- How can information and data from other projects be utilised for the benefit of MH:EK?
- How can information and data from this project be communicated out in an aligned way for lessons learned, and to assist other similar national projects?

Energy modelling process overview

The Milford Haven Energy Kingdom (MH:EK) project aims to make the area a Smart Local Energy System. Part of the project has involved carrying out detailed energy modelling to determine the most optimal solutions to implement to allow the area to progress towards decarbonisation.

The modelling process that has been carried out to determine the best solutions is shown on the process diagram in Figure 3. There are of two key stages that the project has gone through.

Firstly, the determination of a propositions shortlist. This stage involved an initial data gathering process, obtaining high level data on existing and proposed energy systems for the area, collating them and storing in a relational database which was exposed via a browser-based webform.

A system architecture base interface model was produced, and hypothesis testing of certain scenarios was carried out to determine a longlist of propositions. Following the creation of this longlist, all partners carried out a multi-criteria analysis of the highlighted propositions to determine a shortlist to be modelled in more detail.

Once the shortlist had been determined, the second stage of detailed, dynamic energy modelling could progress. This process required a more targeted data gathering stage to obtain more detailed data for the modelling process. The modelling data was then exported from the database and ran through a Python script to produce a series of 'yaml' (model input) files for each proposition. The modelling tool used was 'Calliope' which allows multi-vector analysis that can be weighted depending on the scenario.

Once the modelling was complete, the results were formatted and input into a PowerBI dashboard for aggregation, visualisation and analysis purposes. This was used also to identify data gaps or any further scenarios and/or sensitivity tests that could be run, sometimes requiring further data gathering. Once all scenarios for each proposal have been run, a preferred solution can be recommended for implementation into the energy system.

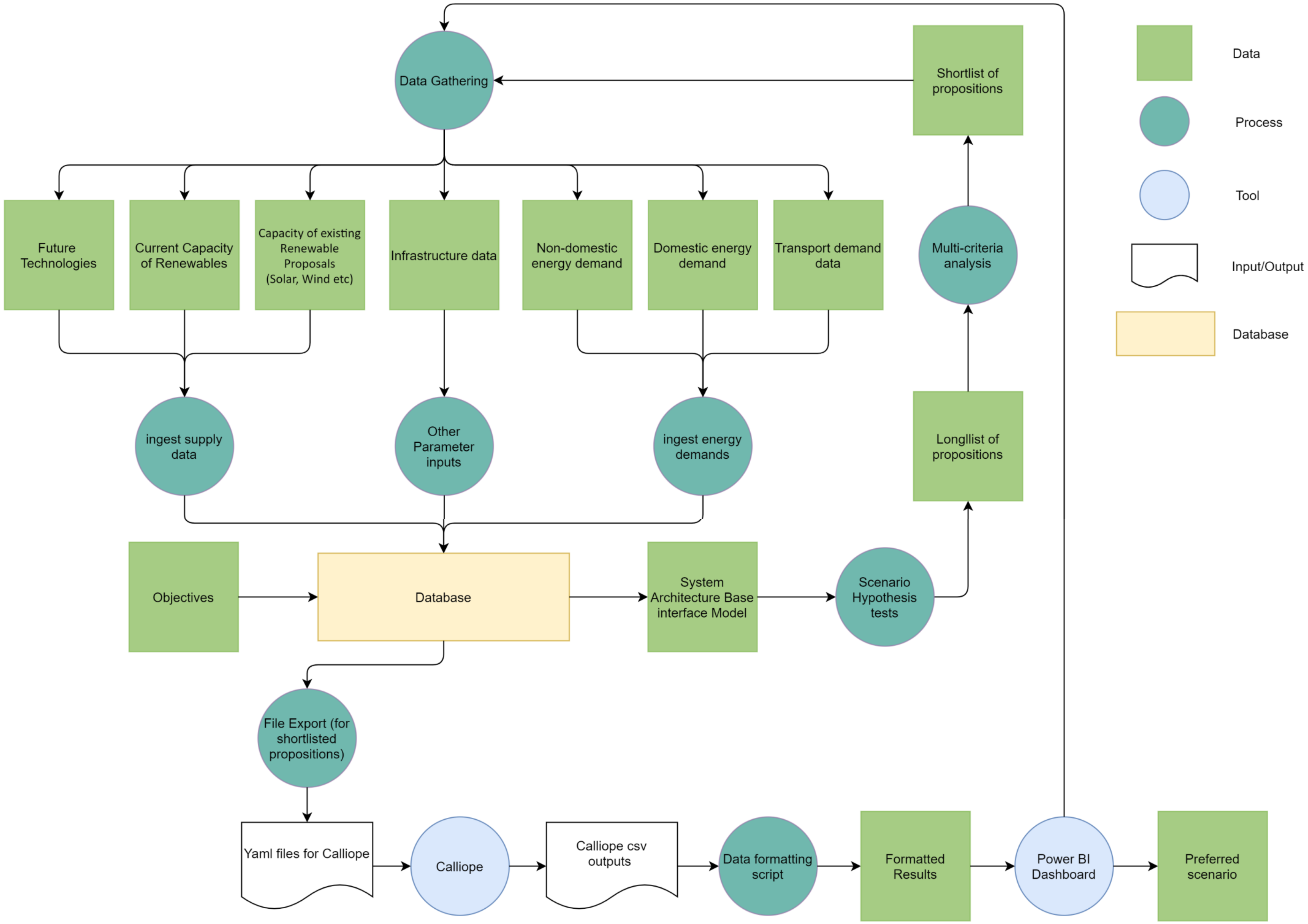
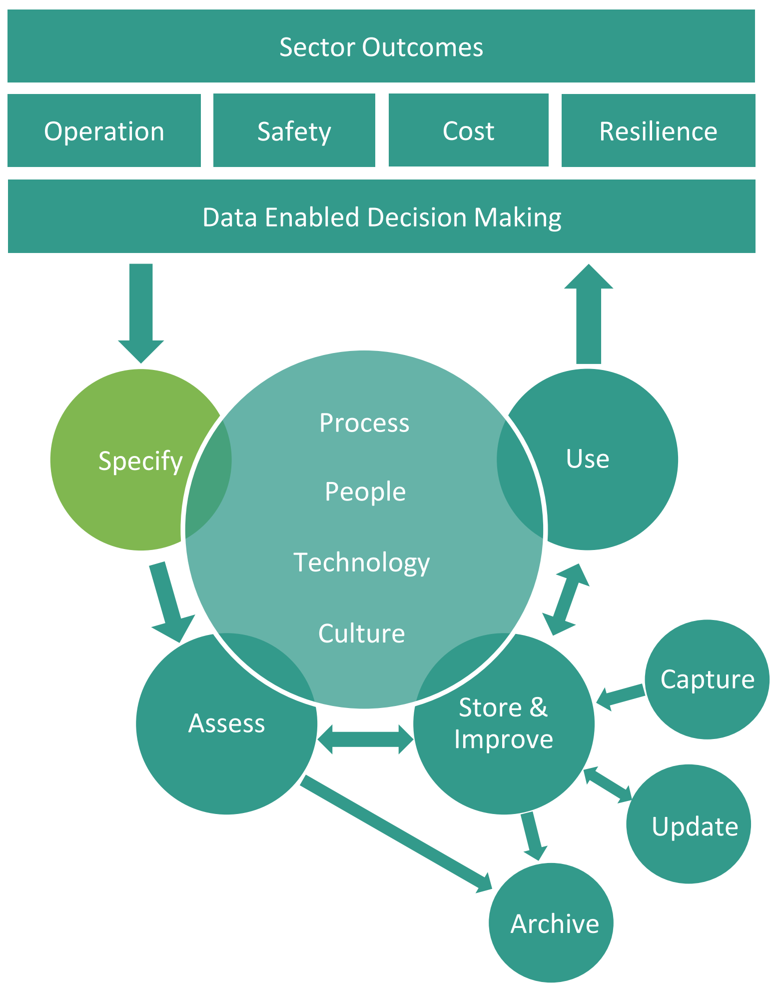


Figure 3: A workflow diagram of the energy modelling process illustrating which datasets were used and when

2. Current state of energy data

What have we experienced on this project?

Energy data lifecycle: Specify



Specify

The specification of data is key to ensure that it is standardised, fit for purpose and fit for future reuse. The correct specification of data allows for effective storage and use of the data, enabling the decision-making process.

The project required both supply and demand energy data for heat, electricity, cooling and transport. Multiple project partners, utility providers and publicly available data sources were used or approached to collate data for the energy modelling exercise.

Data covering both existing and proposed energy usage was required to accurately model the local area's energy system. Data about future propositions was requested from the partners concerning their future developments, and their predicted energy supply or usage.

The modelling process required data to be in a specific format and structure to ensure consistency between inputs. For example, the modelling inputs for the demand of a building required the gas and electricity usage at hourly intervals to accurately track the usage profile.

To ensure that data received was standardised and fit for purpose, spreadsheet templates were developed and issued to data providers to complete. These could be more easily imported into the project data storage systems and ultimately input into the energy modelling.

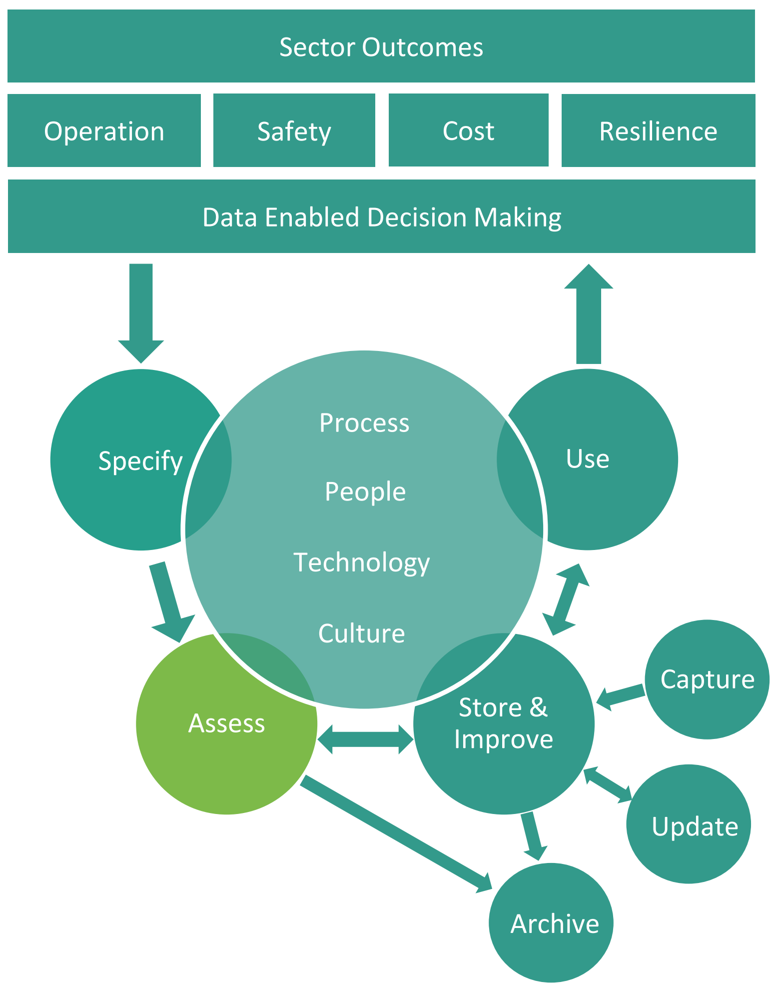
In May 2021, the Welsh Government released and updated 'Welsh Public Sector Net Zero Carbon Reporting Guide' *[6]*. . This guide sets out *'to develop a universal guide set of instructions for use by Welsh public bodies, to estimate baseline emissions, identify priority sources and to monitor progress towards meeting the target collective ambition of a carbon neutral public sector by 2030' [6].* The guide also provides a standardised format for datasets. Pembrokeshire County Council had produced datasets in these standards which they provided for project use.

Challenges

The MH:EK project team implemented some positive processes to assist in specifying data however, some challenges surrounding specification were still encountered:

- Without a clear line of sight to the output and decisions required, it was a challenge to know what data was required for modelling and more importantly, what data to request of data providers. As a result, large quantities of data was initially requested and received that was not key to the detailed modelling exercise.
- Data providers generally do not specify, define and collate datasets for the purposes of energy modelling, and are more likely to produce them for their own organisational purposes and drivers. As a result, existing datasets were not always aligned or formatted to the requirements of the project.
- Where datasets were provided in the Welsh Government Reporting Guide standardised format, the raw data was very valuable and consistent however, as it had previously been specified for carbon reporting and not energy modelling, the data formatting was not fully aligned to the project data requirements.
- Datasets proved challenging to specify. This was found to be prevalent among the more recent technological advancements where there is a lack of practical use cases. Obtaining datasets for these technologies required approaches to be developed such as deriving data from proxy datasets and use cases where similar technologies were implemented.
- Barriers to accessing the datasets which were not publicly available:
    - Payment requirements
    - Strict data agreements
    - Sharing of raw data collected for previous exercises and studies

Energy data lifecycle: Assess



Assess

Once datasets had been received from the data providers, it was assessed for its use and suitability for project purposes. This assessment involved determining if the data required manipulation to be suitable for project storage and use. If the data was unsuitable, and data manipulation would not resolve the issue, alternative or proxy datasets were sourced.

Where alternative datasets were required, publicly available datasets were used where possible. The Department for Business, Energy & Industrial Strategy (BEIS) open datasets were used where applicable, as a trustworthy source across the project to determine energy supply values, usage and cost.

Where publicly available data was also not readily available or suitable, proxy datasets were derived. These proxy datasets are not as accurate as they are based on national profiles and do not accurately represent the local energy supply or demand.

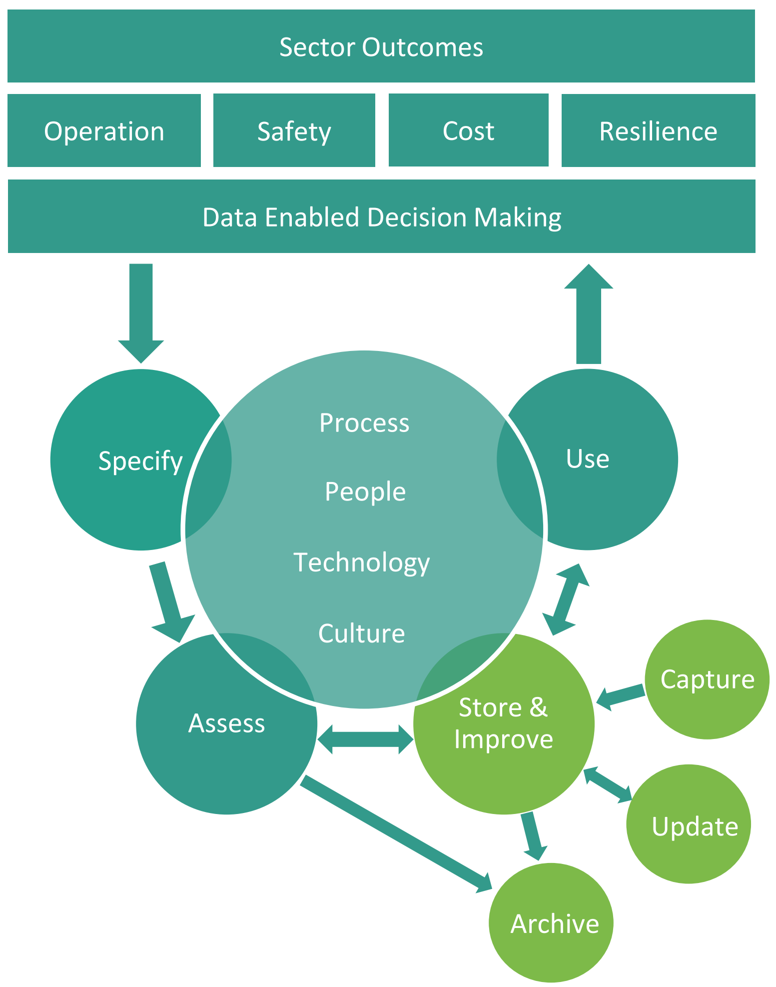
There is no industry standard specifications or documentation for data collation and assessment. The datasets were received in varying formats and included content and metadata that differed between each dataset. As a result of this, each dataset was required to be assessed individually to determine its suitability prior to being input into the project database.

Challenges

Across the project, there were some key challenges with the assessment of datasets and as previously mentioned, the data provided was found to differ in many ways. Some key challenges and differences found when assessing the datasets are highlighted below:

- The following differences in the received datasets caused considerable amounts of manipulation of datasets across the project to ensure it was suitable for the project storage and use.
    - Format of the data – different data providers had inconsistent internal data structures
    - Frequency of the data – data was provided in varying timescales
    - Data content – data received was not always complete, and regularly needed unit conversions
    - Data confidence – varying levels of meter data were provided as validation
    - Spatial and non-spatial data – different storage and formatting requirements depending on data type
    - Security – Some data could not be provided or shared due to GDPR, security processes of the data provider and preventing data misuse
- Publicly available datasets could not always provide the desired accuracy for the energy modelling exercise. Additional external sources were investigated where possible to validate datasets or technological characteristics.
- Reliable sources were often difficult to locate and as a result, led to varied reliability of the data provided.
- Certain datasets were more challenging to receive and implement than others due to restrictions or lack of available data. The data restrictions were one of the key factors that limited the ability to share / open datasets between partners without predetermined data agreements in place.

Energy data lifecycle: Store & improve



Store & improve

Once datasets had been specified and assessed for project use, the datasets can be improved and manipulated where required and stored in the project database.

The project team had identified from an early stage that for the purposes of the modelling exercise, the use of Excel spreadsheets would not provide the ability for machine/scripting access or central data management. A browser-based system utilising an SQL database was produced and adopted across the project as the common data environment and enabled the project team to query and update data within a web browser and produce modelling inputs more efficiently.

The browser-based system also provided a mapping system to store and utilise the spatial data received and allowed the project team to visualise some of the received datasets.

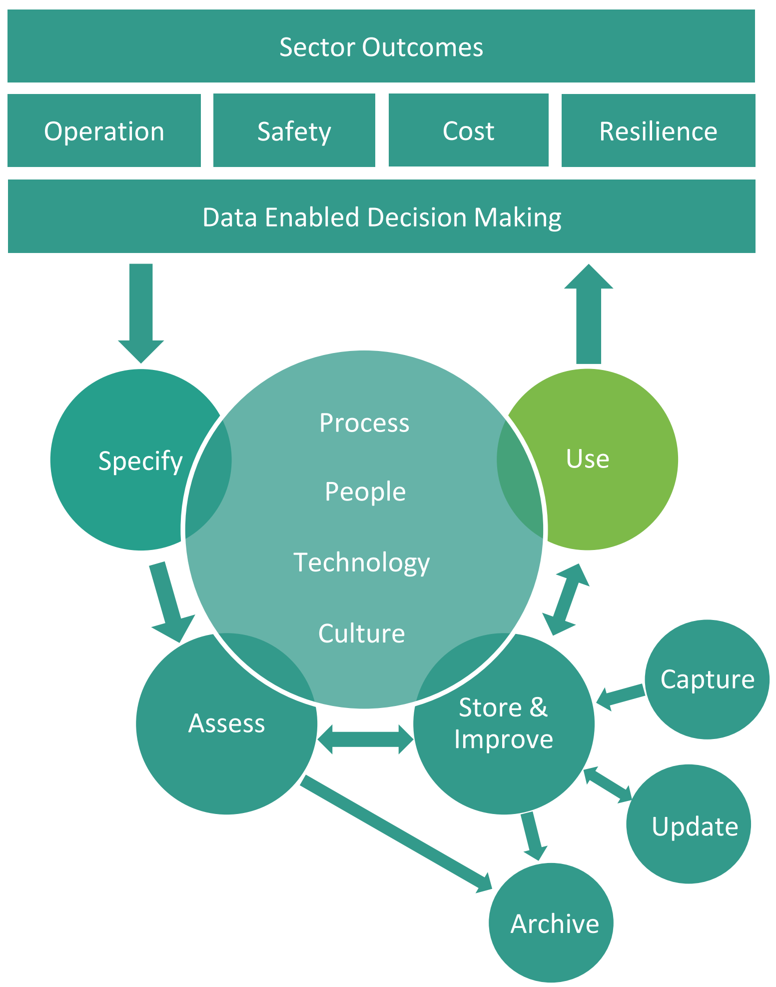
Certain datasets were required to be specifically formatted prior to being input into the database to ensure consistency. As highlighted during the assess phase, considerable assumptions, manipulation and aggregation of datasets was required for project use.

Challenges

Across the project, challenges were encountered surrounding received datasets that were not suitable for immediate project storage or use. The key challenges encountered were as follows:

- Significant manipulation of datasets was required to ensure consistency of data formatting and metadata and enable the usage of the datasets in the energy modelling process. Manipulation had time and cost implications.
- The manipulation process of datasets often required aggregation and assumptions to ensure they included data required to enable the modelling process. These assumptions were more prevalent with data surrounding newer technologies as it is not as open or readily available.
- The common data that needed to be assumed or aggregated included profiled energy supply data and the cost of both the implementation and operation of the systems or technologies.
- Determining and communicating the accuracy and reliability of datasets where assumptions and aggregation was required proved to be challenging. The lack of available sources and use cases made it difficult to determine whether data represented a positive or negative outlook on the proposal.
- The confidentiality and license limitations (limiting third party data sharing) of certain datasets is preventing the browser-based platform from being shared between project partners.
- The core project team were energy specialists looking at the technical modelling process. This team was supported by a team member with core skills in data usage and management of the project database including storing, querying and editing data.

Energy data lifecycle: Use



Use

Once all collected datasets had been stored in the browser-based database and web portal, the data could be queried and exported. The database allowed data to be exported in a format that could be run through automated formatting scripts before being imported into the modelling tool.

As some datasets required assumptions, aggregation and manipulation, there were varying confidence levels which were tracked as data was being collated. Confidence levels were determined and shared with the decision makers throughout the data gathering and modelling process to assist the outputs.

The modelling tool allows for multi-vector optimisation based on select criteria. The modelling output then provides recommendations based on data inputs which are dependent on predetermined scenarios. The scenarios that were run as part of the modelling were agreed with the partners of the project. There is currently no sector standard scenarios to be carried out for energy modelling. Key timescales (Now, 2035, 2050) have been highlighted and modelled to provide a future vision of the proposals and maximise the value of recommendations.

Once the modelling process is complete, an output file is produced and run through an additional automated script to format the results and allow for querying. PowerBI dashboards and graphics have been produced for the visualisation of results.

To assess different scenarios and validate the modelling results, sensitivity analysis of model inputs were run. The analyses allowed assessment of extreme scenarios, and an assessment on the impact severity of each dataset on the outcomes of the modelling.

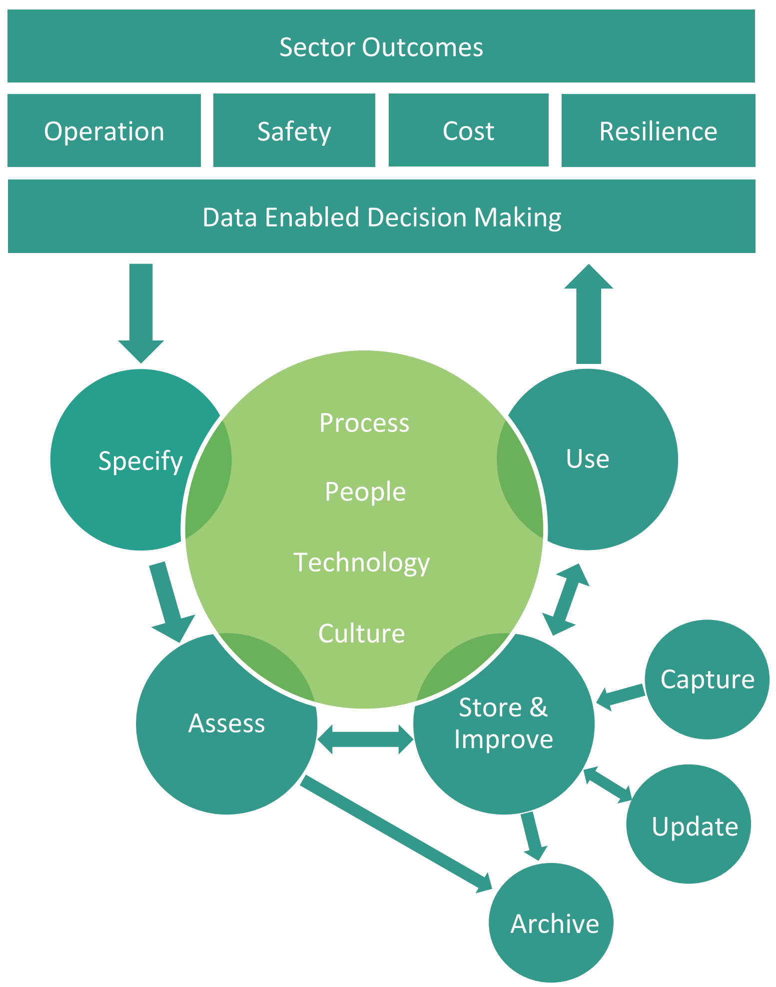
The project outcomes and purposes of modelling have been developed as the project progresses with inputs from different project partners through multicriteria analysis exercises and workshops. This has in turn allowed the most appropriate requirement weighting and scenarios to be run in the modelling.

Challenges

The effective use of data is key to maximise the value of the outcomes of the modelling process. The challenges faced across the project when using the data are as follows:

- A collaborative vision of the required outputs of the modelling process was not initially in place, meaning that from data gathering through to the use of data, the modelling process took a while therefore to be optimised to suit the project needs.
- The determination of confidence levels in datasets was required to enable appropriate and realistic analysis of the modelling outputs. As the received datasets differed in format, accuracy and metadata, defining, categorising and communicating the corresponding confidence in datasets proved challenging.
- The technologies utilised during the project provided an overall benefit by allowing the most beneficial propositions to be determined. However, specialist skills were required to establish the following technologies throughout the project which should be factored in future such endeavors:
  - Use of SQL databases providing a robust server- based solution for storing data
  - Web-based platform – Accessible to the wider team via a browser including web mapping, data forms sharing project information which was editable.
  - Visualisation – Converting the modelling outputs into the PowerBI to present the aggregation of data and filtering by users.
  - Usage of Calliope – Calliope was the main project tool which provided flexibility and a powerful modelling tool.
- Conveying the data / results relevant to the partners and individuals with a non-technical background often required further manipulation and description to assist in answering the project questions of which propositions are 'best'.

Energy data lifecycle: Core enablers



Core enablers

The core enablers are key to ensure the value and efficiency of the energy data lifecycle is maximised. Below is a summary of the use and challenges surrounding core enablers on the project.

Process

- Once clearly defined project outcomes were determined, effective processes for data collation and energy modelling could be implemented.
- Data sharing processes for the purpose of energy modelling were not commonplace in some of the partner or stakeholder organisations, making the collection of valuable datasets challenging.
- There is a lack of sector-wide documentation surrounding the energy modelling processes, but some best data practice guidance is available e.g. from OfGEM *[7]*
- There documentation surrounding processes (data collation and modelling) that have been undertaken on the project is, so far, limited.
- The process by which data providers captured or collated data was not often clear or documented, limiting confidence in the data.

People

- There were not always dedicated individuals within organisations/data providers whose role it was to maintain and organise energy related data.
- Data collation for external partners was generally carried out as an existing function of a team within the organisation, or for other purposes (e.g. Welsh Net Zero Carbon Reporting Standards).
- On this project Arup had assigned a data specialist who was responsible for providing data management assistance for the set up of the central database, Web Platform, GIS web mapping and the various webforms which greatly supported the modelling specialists.

Technology

- The Project established a central project portal called Fuse (which is an Arup tool combining different web technologies). This enabled a central area for both spatial (map based) and non-spatial tabular data which could be filtered and/or edited in browser (if user permissions granted).
- Behind the Fuse Platform, an SQL Server Relational Database was used to better organise and manage the data. The SQL database was found to work well as a database from a data storage and management system perspective however, there was a need for specialist skills on the project to be able to alter/query the data.
- GIS web mapping functionality within the portal aided the visualisation of the project area though it was not used for visualisation of the energy modelling results in this instance.
- Automated Python scripts were used to convert the data export from the web-based database into the modelling input.
- The Calliope tool was used for modelling which allowed for multicriteria analysis where weighting can be altered depending on the solution.

Culture

- On the whole the project team found that the stakeholders on the project understood the purpose of sharing data for the exercise and were willing to do so.
- Data confidentiality and agreements prevented the central project portal being shared with stakeholders and partners.
- The project team were open to the idea that data was important and as a result, ensured specific data management skills were brought onto the project.
- The project team were also open to the use of automation and could see the project benefit of it, implementing it where possible.

Dataset quality and availability

An assessment of the collected datasets used in the modelling process was carried out to determine whether certain datasets presented more of a challenge than others. Table 2 below assesses the data against the availability, suitability and required manipulation of data for project use. Each dataset category has been RAG reviewed with comments detailing the reasoning and background of data collection.

It is clear from the RAG review that datasets with limited availability and/or require the most manipulation are the future technologies, and the domestic energy demand. The reasons for these limitations are however difficult to mitigate due to data security or a lack of information and uncertainty. Therefore, approaches applied during this project to cope with this should be documented applied in future projects. Other datasets which currently have limited availability will hopefully improve in future as the sector moves to more open data with some initiatives underway.

RAG Rating	Data Availability	Data Suitability	Data Manipulation Required
	Not available	Not suitable in the form provided for use	Proxy dataset required
	Available with limiting agreements	Suitable with aggregation	Manipulation required
	Readily available / Open datasource	Immediately suitable for use	No manipulation required

Data Type	Dataset Category	Data Availability	Data Suitability	Data Manipulation Required	Comment
	Future technologies				Consolidated sets of data were generally not available from external sources for newer technologies. Data and resources from previous Arup projects were used to produce more reliable datasets for modelling.
Supply	Current capacity of renewables				Existing data surrounding renewable energy sources (specifically wind and solar) was collated however, required aggregation and manipulation.
	Capacity of current renewable proposals (solar, wind)				Renewable proposals were more challenging to obtain data. There were uncertainties and assumptions made as these proposals had not been finalised or constructed.
	Non-domestic energy demand				A lot of the non-domestic buildings could provide historic heat and electrical data. The data varied between half hourly metred data, and annual totals of energy usage, meaning manipulation was required.
Demand	Domestic energy demand				Due to GDPR, domestic data could not be sourced and needed to be aggregated based on factors of the buildings including building usage, size, age etc which required significant manipulation.
	Transport demand data				Pembrokeshire County Council provided a report detailing their transition to zero and ultra low emission vehicles, this provided appropriate data however required this to extracted from amongst other provided data. This also required manipulation.
Other	Infrastructure data				Existing infrastructure data included both spatial and non-spatial data. Spatial data required more manipulation on the project.

Table 2: Assessment of data against the availability, suitability and required manipulation of data

Challenges – Summary and Comparison

Comparison of challenges faced

A series of sector-wide issues and challenges have been highlighted in the Energy Systems Catapults (ESC) 'Energy Data Taskforce Report' *[5]*. ESC also produced an Energy Revolution Integration Service Insight Paper titled 'Enabling Smart Local Energy Systems: The value of digitalisation and data best practice' *[8]* which included existing challenges of Smart Local Energy Systems (SLES's) across the UK including MH:EK. This section will compare the challenges found in MH:EK with the challenges highlighted within these reports to determine if the project reflects their findings.

Energy Data Taskforce Report

The Energy Data Taskforce Report, identified two key issues within the energy sector surrounding data, and these are 'Data gaps' and 'Extracting value'. It also highlights existing challenges the sector faces to resolve these issues such as fragmentation, power imbalances, culture and skills.

Data gaps – the report notes that in its current state, data quality is often poor, and that leads to it being inaccurate, imprecise or missing. It states that this is the case due to reasons including data existing in non-digital formats, data being collected but not stored, and data just not being collected. The MH:EK project encountered this issue on some occasions when collating datasets for modelling, particularly the lack of consistency in data received, requiring additional manipulation before using. This could be largely down to the data provider collecting data for internal reasons and requirements, and not collecting with a view of sharing it and implementing it into energy modelling exercises.

Extracting value – the report refers to organisations collecting and controlling energy systems data not being able to extract its full value due to highly restricted access, poor data discoverability, strict terms and conditions, and low quality/consistency of data. The MH:EK project mirrored the report findings and found difficulties with data restrictions and strict terms and conditions with data providers across the project. This currently prevents the project from having open data between all partners. The lack of consistency across all datasets received and used on the project also required considerable amounts of manual manipulation of data to make it suitable for storage and modelling.

Fragmentation - The MH:EK project confirmed that it was challenging to collect, manipulate and utilise accurate data when information surrounding costs and benefits are distributed unevenly across many organisations, and there is not one open data source for the sector to use.

Power Imbalance – The MH:EK project is a SLES and is pushing for innovation within the energy sector. The data providers on the project generally collected data for the benefit of the organisation and not necessarily for the 'innovators' to produce energy modelling. A regulated monopoly cooperation as suggested in the Data Taskforce report with incentives to assist innovation in the sector would be beneficial.

Culture: The MH:EK project and partners had a positive culture surrounding innovation, decarbonisation and digitalisation. The data collection however, occasionally reflected the taskforce findings where risk aversion/policy from data providers were restricting the usage and ability to share the data rather than supporting collaborative, data driven solutions.

Skills: MH:EK found data management skills on the project to be challenging but important. The project team identified a requirement for a data specialist on the team, who worked to ensure data storage and management systems were implemented effectively. In the data collection process, the data providers generally did not always have local data management roles coordinating the collection and storage of data. This complicated the data collection process.

Energy Revolution Integration Service Insight Paper *[8]*

The authors of this paper had facilitated a workshop and conducted interviews with a range of stakeholders and organisations including Local Authorities, network owners and operators, technology developers, investors, community groups and the end user. All parties were involved with SLES's, and the outcome was to determine the main challenges they were facing surrounding data and provide recommendations for improvement. An extract of the summary of key challenges taken from the paper is shown below *[8]:*

- *There is difficulty in accessing high granularity energy data, which is required to demonstrate the value that SLES can offer.*
- *The ambiguity of commercial sensitivity and the intellectual property rights of energy data represent a barrier for wide access to data and information.*
- *It is challenging to gather informed consent or to access anonymised personal data to understand how users interact with new energy technologies.*
- *Gathering and combining data from multiple sources is an extremely time-consuming task due to the lack of standards and inadequate data sharing procedures.*

The challenges presented above have all been highlighted as challenges within this project. This also implies that these challenges are being experienced across all stakeholders involved in SLES projects, and so were not specific to Milford Haven.

3. Future state of energy data

What does good look like?

ENERGY KINGDOM
DEYRNAS YNNI

Future of Energy Vision

National Grid ESO – Virtual Energy System *[9]*

National Grid Electricity System Operator (ESO) has recently launched an ambitious plan to create (in collaboration with the wider industry) a real-time digital replica of Great Britain's energy system.

The initial plan is to create a Common Framework for data sharing within the sector, which will be developed through advisory groups and stakeholder engagement. It is important that as one of the more advanced potential SLES, that MH:EK partners contribute to these advisory groups to shape the future plans and standards.

Open Energy Initiative – IceBreaker1 *[10]*

This initiative to modernise energy data access is in its pilot phase. It aims to create an open marketplace for energy data, a mechanism to search and access control and deliver open standards to address cohesion and interoperability of energy data.

This initiative has consulted widely with industry through advisory groups and established a pilot open energy service *[10]* and has published reports and use cases.

National Grid – Shaping the Gas Transmission System of the Future *[11]*

National Grid Gas Transmission Operator (GTO) have embarked on a wide consultation of the gas sector about the future of gas which includes also the inclusion of hydrogen gas within the fuel mix over time and a market plan.

As one of the key regions for the potential production and distribution of hydrogen in the country it is important that the work from this project aligns to the national plans and initiatives.

Energy Networks Association– Gas Goes Green *[12]*

ENA have established the Gas Goes Green initiative in April 2020. It provides an industry wide view of progress on installing the required infrastructure to decrease the reliance on natural gas, replacing it with biomethane and hydrogen using the existing infrastructure.

It has established an Advisory group, made up of representatives from industry (including WWU), academia and policy makers to commence the planning and research phase of this. Whilst not data specific, it is still an initiative to be aware of.

Centre for Digital Built Britain (CDBB) - National Digital Twin Programme *[13]*

The National Digital Twin programme (NDTp) is an initiative to create the building blocks to enable sector based digital twins to form an ecosystem of interconnected digital twins. Through this there will be value created through shared, high quality secure data. This will be a federation of digital twins which requires a common information management framework (IMF), which has been developed by the programme, in order to share data and collaborate.

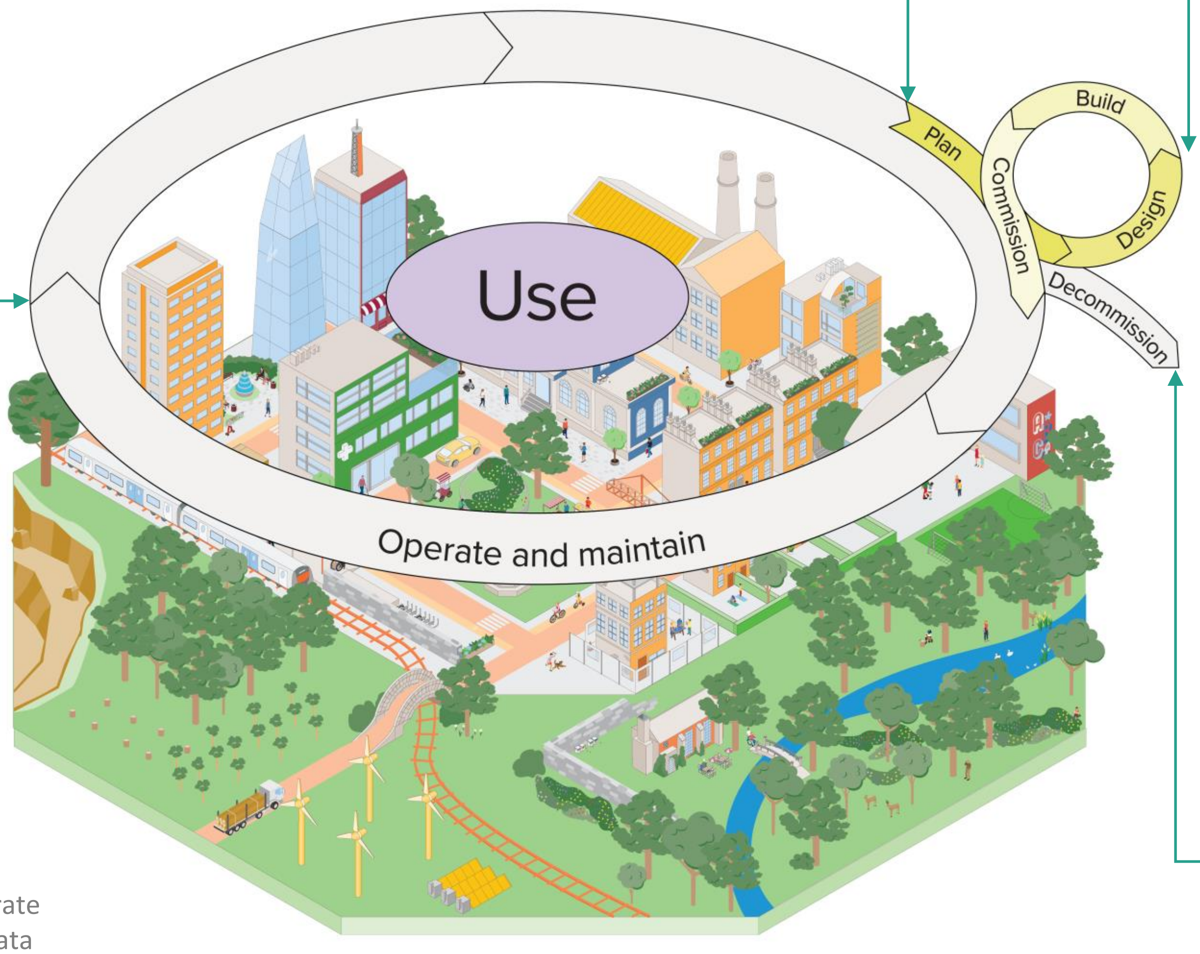
Any energy digital twin in future, and therefore the Virtual Energy System Common Framework and Open Energy will likely align to this national digital twin IMF framework in order to benefit from access to related systems such as transport, telecoms etc.

Energy and Building / Transport Asset Data lifecycle for MH:EK

Planning

MK:EK is currently in the planning phase, whereby we are modelling future propositions and scenarios which should identify the most appropriate investments for the region and sector.

This requires data as inputs such as on infrastructure assets, energy demand and supply both now and into the future to determine whether options are cost effective, safe and operational.



© Centre for Digital Built Britain (CDBB)

Operation and maintenance

Once constructed or retrofitted, a Smart Local Energy System must operate within the wider energy system / market, and therefore must provide data streams for the operation of assets in a safe and efficient way and for national energy balancing. Monitoring and control systems will be in place and will produce data which will be useful to the owner organisation for system optimisation and to others e.g. for network management and decarbonisation reporting. Ideally data about the operations (with appropriate restrictions) should be presumed open and shared and form part of the National Digital Twin and Virtual Energy System.

Maintenance of new systems and infrastructure is also a key requirement, and this is where predictive maintenance scheduling can be defined through modelling of infrastructure operations and condition monitoring information. This reduces the cost of ownership and extends the lifespan of assets.

Design & build

The propositions will eventually become funded projects, some of which will require new construction, networks or modifications to existing infrastructure and buildings. This has both an information need and is a source of data. The best time to obtain important asset data and to install built in sensor arrays for future monitoring is during this phase.

It is therefore important to outline the specification and frequency of data required for future operations of the SLES to maximise the investment and put this into the procurement requirements of the construction process.

Data management and building information modelling are now commonplace during the construction process, and so detailed models and datasets can be delivered from this phase which are crucial for the operations and maintenance phase. These data requirements needs to be included within the contractual documents for them to be costed and delivered however.

Decommission

Some infrastructure will ultimately be retired now or in future due to its high energy demand, intense carbon footprint or its end of life has been reached, which needs to be safely and efficiently demolished, recycled or repurposed.

Again, the change to the base energy system of retiring assets, networks and inefficient buildings need to be accounted for in the overall energy system model and planning for the next phase in the future.

Figure 4: The infrastructure lifecycle in the MH:EK context

Future Data Management Vision figure

Proposed data management structure

The governance of Energy Datasets could be something similar to the figure below developed by the Open Energy Initiative. It relies on data owners capturing and collating datasets in accordance with the Open Energy Data standards and registering those on a central platform. Only data compliant with the standards would be published within the portal. Data can be either shared or open depending on its commercial status.

The data remains stored, and quality controlled by the dataset owner but is searchable through the open energy central portal presenting the data, alongside its metadata. Users interested in accessing that data would register for access which would be governed by a central authority. Once approved they would access the data directly from the providers, potentially through API data services, which could be consumed directly into applications which may be used by others for insight and analysis.

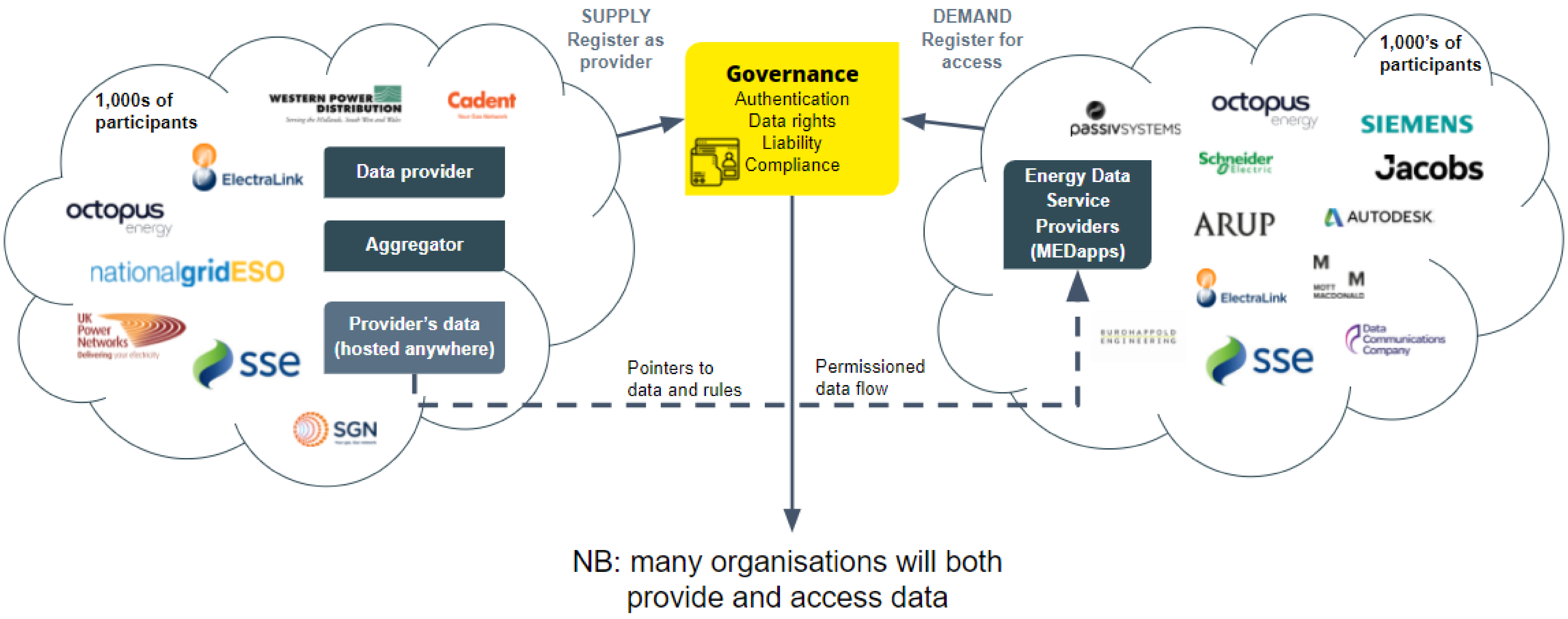


Figure 5: The Open Energy envisaged energy data access model (©IceBreakerOne)

Ideal Future State of Data Ecosystem for MH:EK

Short term

At this stage there has been a focus on the areas where immediate improvements to the ongoing energy modelling could be made based on the findings and lessons learned from the programme to date. This has enabled continuous improvement of the dataset management and analysis required for decision making.

Future state could include:

- A common data portal that is available for all partners to benefit from the central source of truth developed and its visualisation through dashboards and maps more widely than just the modelling teams. This portal has been tailored to meet the needs of decision makers, potentially with the option to adjust parameters in order to visualise the impact of proposed initiatives.
- Improved documentation of datasets within a programme data catalogue – outlining the value and use of datasets in the energy modelling process, and the current manipulation required
- Continuous improvements to modelling have been identified throughout the project and there is be a series of recommendations available which should be followed for future exercises
- A data management working group has been formed within the MK:EK programme partners to ensure that data is provided in a timely and suitable manner, and feedback is received to clarify data needs in future. A representative of this group is part of the NG ESO Virtual Energy System Advisory Groups and Shaping the Future of Gas groups and contributing to these country wide initiatives to improve data sharing.

Medium term

At this stage certain propositions will have been selected for investment and will now be entering the phase of seeking investment then design and construction which will involve a procurement process which sets out requirements of project delivery. This will likely include monitoring and control systems, which could provide vital datasets and asset information for operations and maintenance which are best derived from this stage.

Future state could include:

- Data requirements are documented and available to future projects to comply with. These project data requirements feed into procurement documents and ultimately contracts. This provides the sector with the asset and system data it will need for operation, maintenance and monitoring to meet the net zero target.
- The monitoring systems and ‘right time’ data feeds will be designed with a view to meet both operational and construction needs, but also to meet the needs for energy reporting and access protocols. This will have followed the ESC Data Taskforce recommendations to the sector for open, coordinated and digitalised data being available.
- Ongoing modelling and planning for additional reductions of carbon in the production of energy/transportation, and changes in energy demand will be required as different investments progress. There has been an ongoing data management strategy to incorporate system changes into models periodically to reflect the current situation. This has been achieved through cooperation and data standards having been established and followed from the outset of projects.
- Open data sharing is in place for all new investments and from existing energy providers which is captured at source and in accordance with industry standards as set out by the Virtual Energy System Common Framework. This will provide energy usage and system access at agreed frequency in a secure and interoperable format.

Long term

Many zero carbon emitting energy investments are now in full operation, with others in the pipeline building on the success of the MH:EK programme. The key now is ensuring that the energy system is functioning and balanced, and to determine whether the expected energy / decarbonisation forecasts and commercial expectations are becoming a reality. This required the various energy models and analysis to be updated with as built/ as operated data delivered by projects.

Future state could include:

- Operational energy supply and demand data is readily available via open data services, which can be consumed and analysed by authorised organisations for modelling or planning.
- Datasets are fully documented in the form of a data catalogue providing metadata about their source, update process and key attributes meeting international standards, available via the Open Energy website.
- Web based platforms are available to visualise and analyse datasets including maps, dashboards with modelling output scenarios generated on presentation of key input figures.
- Smart Local Energy System data and information is integrated into the National Virtual Energy System digital twin, and the National Digital Twin framework enabling planning on a national scale to determine how best to balance zero carbon energy supply and demand, as well as respond to emergency incidents rapidly.
- Data Governance is practiced across multiple data owners, into which new and existing energy producers and suppliers cooperate and comply with standards and contribute to the decisions on any changes.
- Asset maintainers have all the data they require to plan interventions through predictive modelling of asset performance and planned outages for maintenance which will better inform other energy system operators.

4. Recommendations and roadmap

How do we get there?

ENERGY KINGDOM
DEYRNAS YNNI

Recommendations – Sector wide

Introduction

The recommendations made within the Data Taskforce Report largely meet what is required within the Milford Haven SLES in future. It is critical that the local system is connected within the wider system, and to do so needs to match the data standards and sharing protocols. This section presents those Taskforce recommendations, and the next provides more detail specifically for MH:EK to consider.

Energy Data Taskforce recommendations

The Energy Systems Catapult Data Taskforce outlines the following actions which will address the 5 main recommendations. Detailed steps and their relative timescales and resolution were also put into a chart in the report, on top of which the short-, medium- and long-term actions for this programme have been added (Figure 6):

- **Digitalisation of the Energy System**: new legislative and regulatory measures to drive the capture of new data, improving existing data and developing 'Digitalisation Strategies'
- **Maximising the Value of Data:** new legislative and regulatory measures to ensure data is Presumed Open, promoting the development of 'Structures, Interfaces and Standards'
- **Visibility of Data:** developing a Data Catalogue to improve data visibility
- **Coordination of Asset Registration:** developing a new asset registration strategy
- **Visibility of Infrastructure and Assets:** developing a Digital System Map to increase visibility of assets and promote investment and new markets

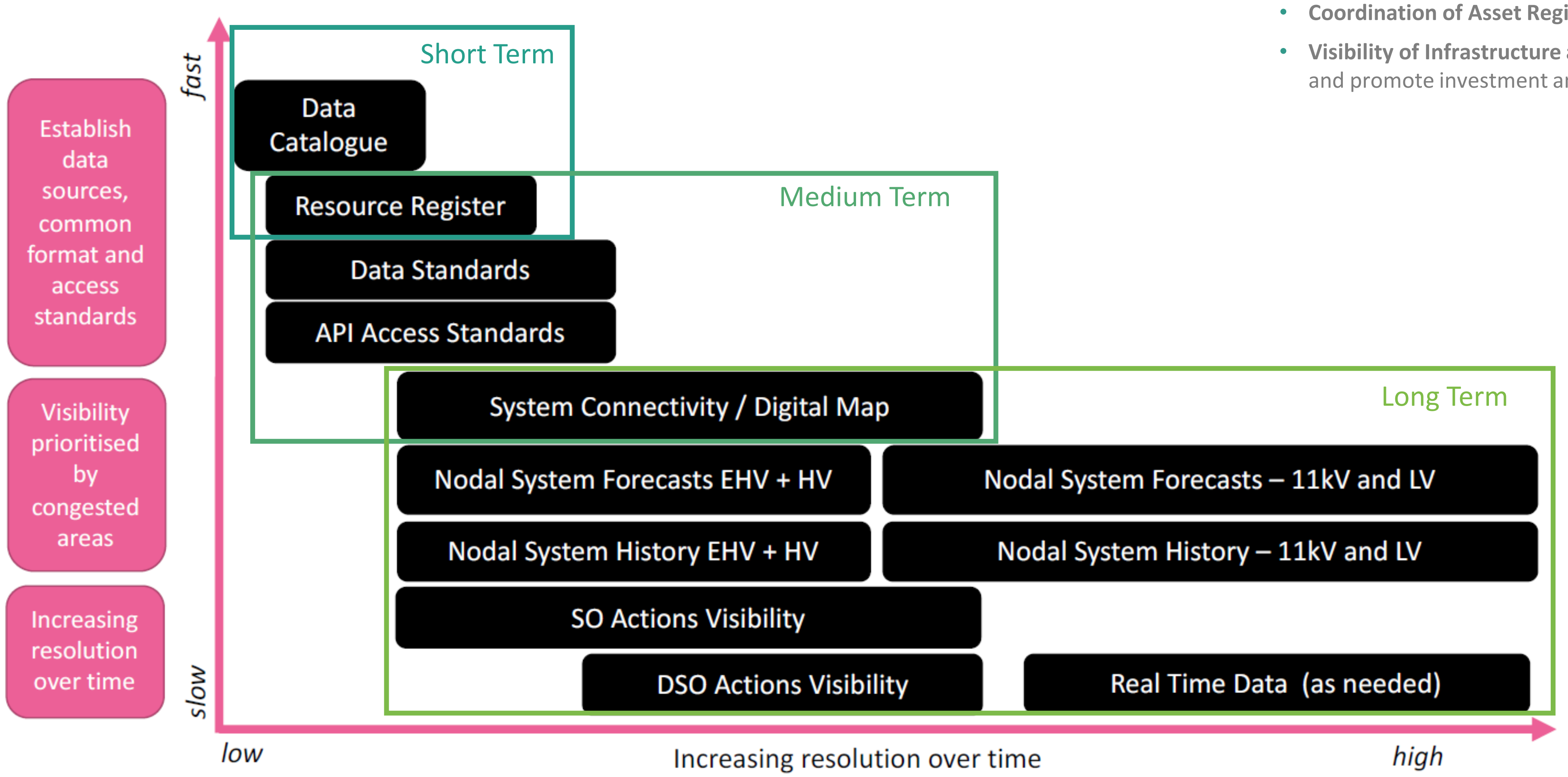


Figure 6: Recommendations made by the ESC Data Taskforce, overlaid by a timescale considered feasible for the MH:EK SLES (©Source: Figure from "Data for Multi-Party System Operation: Energy Data Taskforce Appendix 5", by Energy Systems Catapult)

Recommendations – MH:EK SLES

The recommendations highlighted in the Energy Data Taskforce Report *[5]* refer to the implementation of steps and solutions across the sector to achieve digitalisation and decarbonisation. As these recommendations are currently not in place and will not likely be in place in the near future, there are further recommendations that can be undertaken on a localised basis to enable and prepare projects like MH:EK and other SLES's ahead of national standards and guidance being implemented. The following recommendations have been formed looking at the sector-wide recommendations whilst considering the current situation of data and challenges within the sector, including what was found on MH:EK and other SLES's.

#	Recommendation	Description	Action Owner	Lifecycle Aspect(s)	Difficulty of Action	Action Impact
1	Common Energy Modelled Data Portal	A common energy model data portal available for all partners to benefit from the central source of truth developed and its visualisation through dashboards and maps more widely than just the modelling teams. This portal could also potentially provide a secure API data service for approved parties to bring data directly into their own systems. The data service link and its metadata, could potentially be referenced within the Open Energy data demonstration portal and would need to be hosted by one of the MH:EK organisations.	Project Team	Use, Store & Improve	2	6
2	Programme Data Catalogue	A catalogue will allow for improved documentation of datasets whilst outlining the value and use of datasets in the energy modelling process. This will also include any datasets which have been manipulated or derived by the modelling team. This will help identify the source, key metadata, details/process of any current manipulation required, and any known gaps and inconsistencies within the dataset. *Data should be managed and shared by the owner of the dataset, and the metadata provides a link to that, and additional information required.*	Project Team (Arup)	Specify	4	4
3	Adoption of a Data Openness Triage process to achieve 'Presumed Open'	All partners adopting an Openness Triage process which considers a range of risk factors and develops an appropriate range of mitigation mechanisms. Ensure all project data that can be open is open (within reason) to enable data and information sharing across the sector. The adoption of this on a sector-wide basis will also allow data sharing and improvements between other SLES projects. This data sharing is more likely to occur at a higher level than the programme, and so the action will likely lie with the Welsh Government and/or OfGEM but the programme can prepare itself for this scenario by setting up the appropriate data agreements and preparing datasets to be open.	Dataset Owners	Specify, store & improve	8	7
4	Formation of a Milford Haven Energy System data management working group	A working group within the MH:EK programme partners to ensure that data is provided in a timely and suitable manner, and feedback is received to clarify data needs in future. A representative of this group should be part of the NG ESO Virtual Energy System Advisory groups, and also part of the NG Shaping the Future of Gas Transmission initiatives, in order to influence future frameworks. Once in place this group will promote the data standards/frameworks in future. This group could have an active role in the audit and verification of system or infrastructure compliance with standards in future in procurement data requirements and deliverables.	MH:EK Partners (PCC Chair)	Specify, Assess, Process	5	6

Table 3: Summary of recommendations for the MH:EK SLES

Recommendations – MH:EK SLES (Continued)

#	Recommendation	Description	Action Owner	Lifecycle Aspect(s)	Difficulty of Action	Action Impact
5	Creation and implementation of an ongoing data management strategy to incorporate system changes into modelling	A data management strategy will support ongoing modelling and planning for additional reductions of carbon in the production of energy/transportation, and changes in energy demand will be required as different investments progress. This can be achieved through cooperation between partners and implementation of data standards having been established and followed from the outset of projects. This will support continuous improvements and efficiencies in the current modelling process.	Project Partners	Use, Process, Culture	4	6
6	Open web based visualisation platform	Production of an open visualisation platform which will allow users to understand how the system is configured and how they are connected to it for the Milford Haven SLES area (or possibly Wales). Staged implementation can be considered as the platform can represent the systems normal running configuration, or a smart system can be implemented which incorporates real-time running arrangements. Examples of existing implemented platforms are GLA's FlexLondon which looked to find synergies between existing and new flexible energy developments, and the Australia Renewable Energy Mapping Infrastructure which was produced to enable renewable energy projects across the country.	Project Partners	Use	7	4
7	Data requirements for future projects	Defining data requirements at the procurement stage of design and construction of future projects will provide the sector with the asset and system data it will need for operation, maintenance and monitoring to meet the net zero target. These will feed into the construction contractual documents to ensure that projects deliver the necessary data. This could come from the NG ESO initiative to create a Virtual Energy System and a Common Framework.	Project Partners	Specify, Process	3	4
8	National Digital Twin / Virtual Energy System integration	Ensuring data alignment along the Gemini principles, allowing the project to be incorporated into the National Digital Twin as it is progressing. This would be best done by being part of the NG ESO Virtual Energy System initiative which seeks to create a Common Framework for sharing data between various stakeholders. MH:EK could be represented within the Advisory Groups to have their input based on the experience of this project.	Sector Working Groups	Specify, Use	8	6
9	Contribution and adoption of national energy data standards and access protocols	The Milford Haven: Energy Kingdom programme being one of the first such Smart Local Energy System projects should play a key role in the setting and adoption of the data standards which have been also recommended by the Energy Systems Catapult Data Taskforce. This would be achieved through representation of the MH IM Working Group at key standards committees if formed out of the Virtual Energy System initiative for example. Once the data standards have been agreed, moving existing datasets into such standards will require a change process from source through to derived datasets which will again require active participation and involvement. Enforcement and governance of the standards however is likely to be a regulatory requirement, and therefore be outside the authority of this group, but support could still be provided.	Project Partners	Specify	4	7

Table 3: Summary of recommendations for the MH:EK SLES (cont.)

Roadmap

#	Recommendation
1	Common Energy Modelled Data Portal
2	Programme Data Catalogue
3	Adoption of a Data Openness Triage process to achieve 'Presumed Open'
4	Formation of a Milford Haven Energy System data management working group
5	Creation and implementation of an ongoing data management strategy to incorporate system changes into modelling
6	Open web based visualisation platform
7	Data requirements for future projects
8	National Digital Twin integration
9	Contribution and adoption of national energy data standards and access protocols

Table 4: Summary list of recommendations

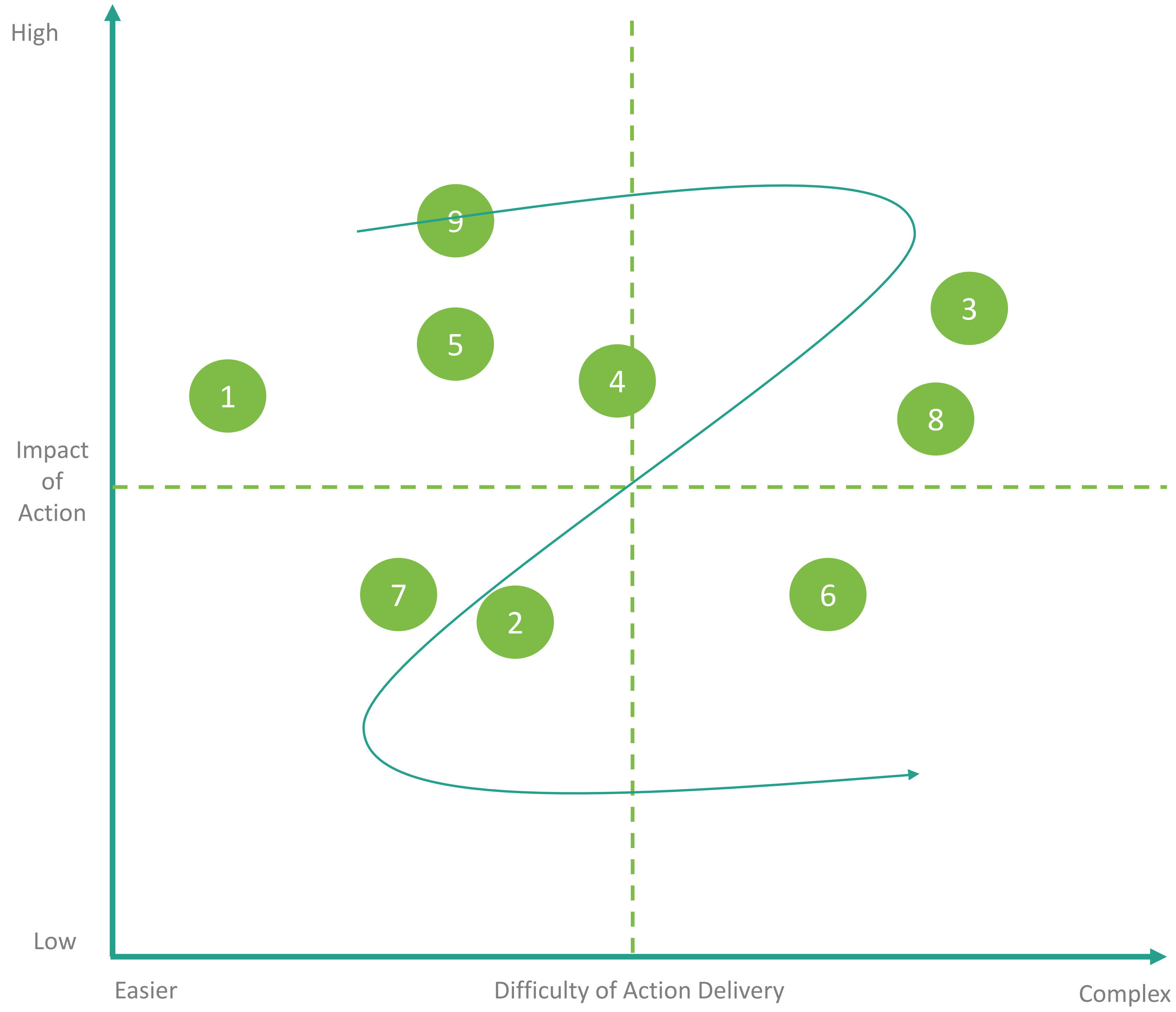


Figure 7: Potential roadmap of implementation based on complexity and impact of recommendation

Recommendations – At different scales

Introduction

The need for the recommended data and actions will differ depending on what is possible at different organisational scales, and the influence they will have. Certain actions will be required at different levels of the sector as outlined below in Figure 8.

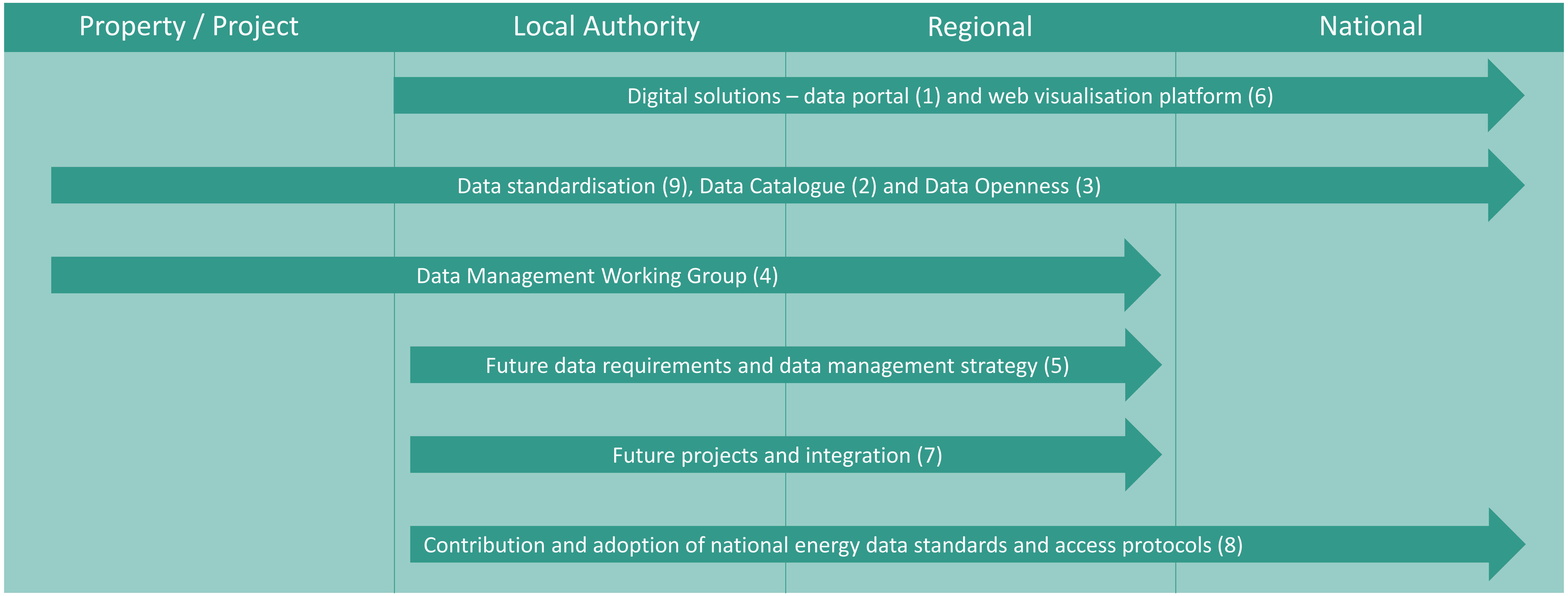
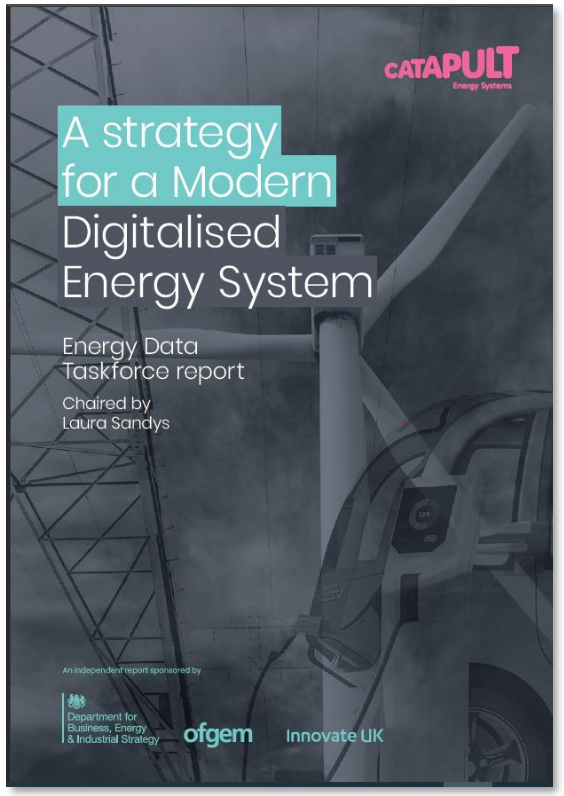


Figure 8: The likely organisational scale which the recommendations if made will impact or involve

5. Conclusions

ENERGY KINGDOM
DEYRNAS YNNI

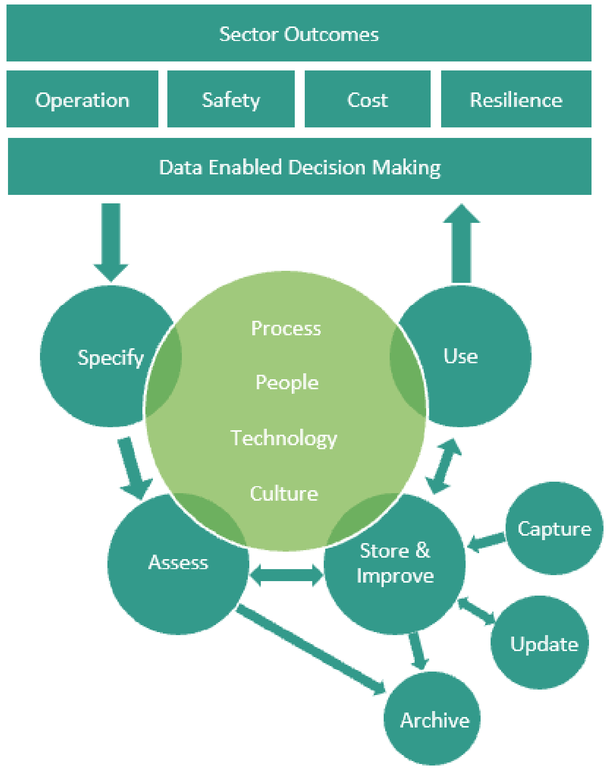
Conclusions



Aligning to the recommendations made in the Energy Data Taskforce report *[5]*

Many of the recommendations made within this report are highly relevant to the programme, and activities should be aligned to these now and into the future. This includes creation of data catalogues of derived datasets, adoption of a national data standard and other initiatives such as system maps. Some of these will likely be outside the control of the programme but as they come into place in future, MH:EK will benefit from adopting and enforcing them, to become a critical part of the national progress towards decarbonizing the national energy system.

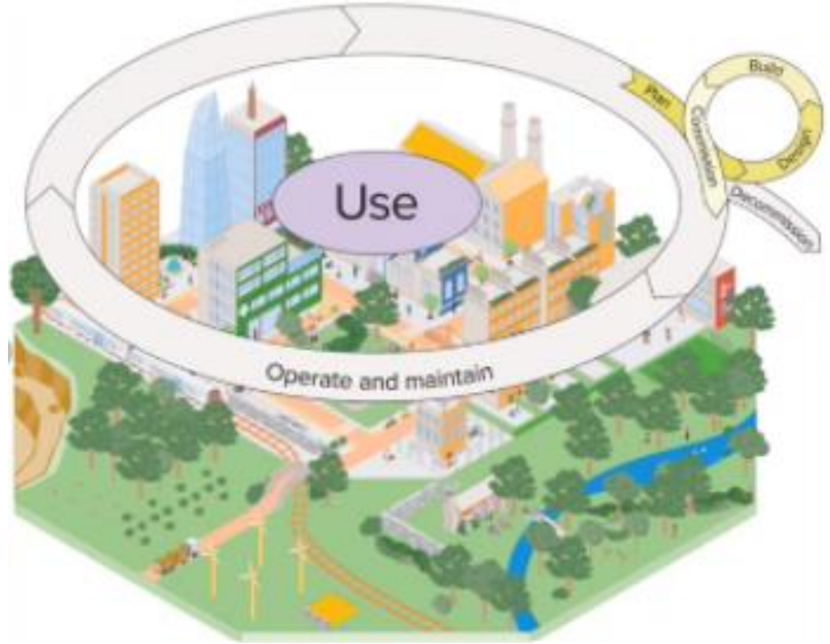
Other key recommendations to work with or adopt initiatives such as Open Energy and the Virtual Energy System which will benefit many organisations both today and into the future, enabling innovation and effective management and integration of the smart local system in the national system. Benefits such as the export or import of renewable energy become possible if data is well documented, consistently structured and readily available.



Focus on the energy data lifecycle, and especially the core enablers required to achieve the programme outcomes

Some of the main recommendations made in this document are related to some of the core enablers to data management. Most importantly that there is an information management working group established with data managers from relevant organisations for the MH:EK programme to take collective ownership and governance of the data ecosystem required.

It is important that the appropriate processes (including policies) have been established for future modelling and reporting to determine if the propositions provide the sector outcomes of a decarbonised and decentralised energy system. A culture also should be established that open and standardised data is an important enabler for the development and success of a smart local energy system and its integration into the wider network. It is critical to ensure this aspect is not overlooked in future projects.



MH:EK should consider the full Information Lifecycle of the infrastructure

This programme is more than an energy modelling exercise as it will be recommending propositions for investment. Therefore, the data ecosystem roadmap should consider the full asset lifecycle past modelling into design / construction and Asset Management which will each have their own data requirements. The elements such as the specification of data requirements should be considered for the procurement phase for example.

Only by continuing to provide open datasets during the lifecycle of the asset (especially in operation) will the full benefits of a connected digital twin be realised nationally.

References

1. UK Government, *UK becomes first major economy to pass net zero emissions law*, (2019). Accessed at: https://www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law.
2. UK Government, *UK enshrines new target in law to slash emissions by 78% by 2035*, (2021). Accessed at: https://www.gov.uk/government/news/uk-enshrines-new-target-in-law-to-slash-emissions-by-78-by-2035
3. Energy Data Taskforce, *Energy Data Taskforce*. Accessed at: https://es.catapult.org.uk/case-study/energy-data-taskforce/
4. Energy Data Taskforce*, Energy Data Taskforce: A Strategy for a Modern Digitalised Energy System*. Accessed at: https://es.catapult.org.uk/report/energy-data-taskforce-report/
5. Energy Data Taskforce, *A strategy for a Modern Digitalised Energy System*, (2019)
6. Welsh Government, *Welsh Public Sector Net Zero Carbon Reporting Guide,* (2021)
7. Ofgem, *Data Best Practice Guidance*, (2021) https://www.ofgem.gov.uk/sites/default/files/2021-11/Data_Best_Practice_Guidance_v1.pdf
8. Energy Systems Catapult, *Enabling Smart Local Energy Systems: The value of digitalisation and data best practice.* Accessed at: https://es.catapult.org.uk/report/enabling-smart-local-energy-systems-the-value-of-digitalisation-and-data-best-practice/
9. National Grid ESO, *Virtual Energy System*. Accessed at: https://www.nationalgrideso.com/virtual-energy-system
10. Open Energy, *Discover energy data*. Accessed at: https://openenergy.org.uk/
11. National grid, *Shaping the gas transmission system of the future,* (2019). Accessed at: https://www.nationalgrid.com/uk/gas-transmission/document/125911/download
12. Energy networks Association, *Gas Goes Green.* Accessed at: https://www.energynetworks.org/creating-tomorrows-networks/gas-goes-green
13. University of Cambridge Centre for Digital Built Britain, *National Digital Twin Programme*. Accessed at: https://www.cdbb.cam.ac.uk/what-we-do/national-digital-twin-programme
14. National Grid, Future of Gas – GMaP. Accessed at: https://www.nationalgrid.com/uk/gas-transmission/future-of-gas